

ENERGY-PRESERVING INTEGRATORS APPLIED TO NONHOLONOMIC SYSTEMS

ELENA CELLEDONI¹, MARTA FARRÉ PUIGGALÍ², EIRIK HOEL HØISETH¹
DAVID MARTÍN DE DIEGO²

ABSTRACT. We introduce energy-preserving integrators for nonholonomic mechanical systems. We will see that the nonholonomic dynamics is completely determined by a triple $(\mathcal{D}^*, \Pi, \mathcal{H})$, where \mathcal{D}^* is the dual of the vector bundle determined by the nonholonomic constraints, Π is an almost-Poisson bracket (the nonholonomic bracket) and $\mathcal{H} : \mathcal{D}^* \rightarrow \mathbb{R}$ is a Hamiltonian function. For this triple, we can apply energy-preserving integrators, in particular, we show that discrete gradients can be used in the numerical integration of nonholonomic dynamics. By construction, we achieve preservation of the constraints and of the energy of the nonholonomic system. Moreover, to facilitate their applicability to complex systems which cannot be easily transformed into the aforementioned almost-Poisson form, we rewrite our integrators using just the initial information of the nonholonomic system. The derived procedures are tested on several examples: A chaotic quartic nonholonomic mechanical system, the Chaplygin sleigh system, the Suslov problem and a continuous gearbox driven by an asymmetric pendulum. Their performance is compared with other standard methods in nonholonomic dynamics, and their merits verified in practice.

1. INTRODUCTION

Geometric integrators are numerical methods for differential equations which preserve structural properties such as constants of the motion, symplectic or Poisson structures, phase-space volume, different symmetries of the system or isospectrality. Preservation of structural properties is often desirable to achieve correct qualitative behaviour and long time stability [19, 25, 33].

In this paper, we address the construction of geometric integrators for mechanical systems subjected to nonholonomic constraints. There is considerable interest in the study of nonholonomic systems since nonholonomic constraints are present in a great variety of mechanical systems in engineering and robotics. For instance, they describe the dynamics of wheeled vehicles, manipulation devices and locomotion systems (see [3, 4, 9, 8, 28, 29] and references therein).

In the unconstrained case, or when the constraints are holonomic, mechanical systems have many distinguishing geometric features. Among the most important are the preservation of energy, the symplectic form constructed from the Lagrangian (Poincaré-Cartan 2-form) and the momentum map in the presence of symmetries according to the Noether theorem. As we will see, when we are dealing with nonholonomic constraints this symplectic form is no longer preserved, and the momentum map is not in general conserved in the presence of symmetries. However, the energy is still a conservation law for the system in the case of linear constraints. We therefore focus our attention on the exact preservation of energy, using geometric integrators, while writing the equations of motion in a format which ensures the nonholonomic constraints are satisfied.

The proposed approach is different from other recent approaches such as [9, 11, 12, 15, 23, 24] where the authors have introduced numerical integrators for nonholonomic systems with very good energy behavior, and properties such as the preservation of the discrete nonholonomic momentum map.

The ingredients of the theory of nonholonomic mechanics are a Lagrangian function $L : TQ \rightarrow \mathbb{R}$ of mechanical type, that is, kinetic minus potential energy, and a vector subbundle $\tau_{\mathcal{D}} : \mathcal{D} \rightarrow Q$ of TQ . This vector subbundle determines the nonholonomic constraints, as we will see in the next section. To develop integrators we first introduce a Hamiltonian description of nonholonomic mechanics in terms of an almost-Poisson bracket. Using the Riemannian metric determined by the kinetic energy, and the standard symplectic structure on T^*Q , we can induce a linear almost-Poisson structure Π on the dual bundle $\pi_{\mathcal{D}} : \mathcal{D}^* \rightarrow Q$. This so-called nonholonomic bracket is isomorphic to the nonholonomic bracket considered in [34].

Now, working on the “Hamiltonian system” determined by the triple given by (i) \mathcal{D}^* as new phase space, (ii) the almost-Poisson bracket Π and (iii) the induced Hamiltonian function $\mathcal{H} : \mathcal{D}^* \rightarrow \mathbb{R}$, we apply energy-preserving integrators to simulate its dynamics. This is a coherent approach since the unique generic quantity preserved by the flow of the corresponding Hamiltonian vector field to \mathcal{H} is precisely the Hamiltonian function. The resulting integrators preserve by construction both the energy and nonholonomic constraints.

To approximate the solution while preserving the energy of the initial nonholonomic problem we utilize a class of geometric integrators called discrete gradient methods. Consider an ODE which can be written in skew-gradient form, i.e. $\dot{x} = \Pi(x)\nabla\mathcal{H}(x)$ with $x \in \mathbb{R}^N$ and $\Pi(x)$ a skew-symmetric matrix. In [26] it is shown that any ODE with a generic first integral \mathcal{H} can be put into skew-gradient form. For a generalisation of these ideas to the case where the configuration space is a Lie group or a homogeneous manifold see [5].

Now, discrete gradient methods are based on the following construction: Let $x \approx x(nh)$ and $x' \approx x((n+1)h)$. Using a discrete gradient $\bar{\nabla}\mathcal{H}(x, x')$, which is an appropriate approximation of the gradient of \mathcal{H} (see Section 3 for details), it is possible to define a class of integrators

$$\frac{x' - x}{h} = \tilde{\Pi}(x, x')\bar{\nabla}\mathcal{H}(x, x'),$$

which preserve the first integral \mathcal{H} exactly, i.e. $\mathcal{H}(x) = \mathcal{H}(x')$. Here $\tilde{\Pi}(x, x')$ is a skew-symmetric matrix approximating $\Pi(x)$. It can be shown that, in \mathbb{R}^n , any first integral-preserving (direct) integrator can be written as a discrete gradient method [17, 30, 31, 32].

For a given nonholonomic mechanical system, the equations of motion in canonical coordinates are generally assumed known. A potential obstacle in applying a discrete gradient method directly to the adapted coordinate system is the need for the user to analytically derive these equations. For this reason, we propose a reformulation of the methods using just information from the original system in canonical coordinates. With this approach the analytic reformulation of the system in adapted coordinates is avoided.

The outline of the paper is as follows. In the next section we will describe the geometric framework for nonholonomic mechanics. The main objective is describing its dynamics as a Hamiltonian system on a vector bundle equipped with an almost-Poisson bracket. The resulting equations of motion in adapted coordinates are seen to be explicitly given in skew-gradient form. In Section 3 we apply discrete gradient integrators to the derived formulation to get energy-preserving integrators for nonholonomic systems. We then rewrite these integrators in an equivalent form by using only the information from the original nonholonomic system. Finally, in Section 4, we verify the properties and the performance of our integration techniques, applying them to several interesting examples: The chaotic quartic nonholonomic mechanical system, the Chaplygin sleigh system, the Suslov problem and the continuous gearbox driven by an asymmetric pendulum. Our methods are compared with other well known numerical methods for nonholonomic mechanics.

2. NONHOLONOMIC SYSTEMS

A nonholonomic system is a mechanical system with external constraints on the velocities [3, 4, 9]. We will only consider linear velocity constraints, since this is the case in most examples. Linear velocity constraints are constraints that are specified by a regular C^∞ -distribution \mathcal{D} on the configuration manifold Q , or equivalently, by a vector subbundle $\tau_{\mathcal{D}} : \mathcal{D} \rightarrow Q$ of TQ with canonical inclusion $i_{\mathcal{D}} : \mathcal{D} \hookrightarrow TQ$. Therefore, we will say that a curve $\gamma : I \subseteq \mathbb{R} \rightarrow Q$ satisfies the constraints given by \mathcal{D} if

$$(1) \quad \gamma'(t) = \frac{d\gamma}{dt}(t) \in \mathcal{D}_{\gamma(t)} \text{ for all } t \in I.$$

We say that \mathcal{D} is holonomic if \mathcal{D} is integrable or involutive, that is, for any vector fields $X, Y \in \mathfrak{X}(Q)$ taking values on \mathcal{D} , it holds that the vector field $[X, Y]$ also takes values on \mathcal{D} . A regular linear velocity constraint \mathcal{D} is nonholonomic if it is not holonomic. Observe that in the case of holonomic constraints all the curves through a point $q \in Q$ satisfying the constraints must lie on the maximal integral manifold for \mathcal{D} through q .

Let $\dim Q = n$. Locally if (q^i) , $1 \leq i \leq n$ are coordinates on Q and (q^i, \dot{q}^i) are the induced coordinates on TQ , the linear constraints are written as

$$\mu_i^\alpha(q) \dot{q}^i = 0, \quad m+1 \leq \alpha \leq n,$$

where $\text{rank}(\mathcal{D}) = m \leq n$. The annihilator \mathcal{D}° is locally given by

$$\mathcal{D}^\circ = \text{span} \{ \mu^\alpha = \mu_i^\alpha(q) dq^i; \quad m+1 \leq \alpha \leq n \},$$

where the 1-forms μ^α are independent. Equivalently, we can find independent vector fields $\{X_a\}$, $1 \leq a \leq m$ such that

$$\mathcal{D}_q = \text{span}\{X_a\}.$$

Observe that $\mu^\alpha(X_a) = 0$, for all $m+1 \leq \alpha \leq n$ and $1 \leq a \leq m$.

Example (Rolling disk). One of the simplest examples of a nonholonomic system is the unicycle, that is, a disk of radius r which rolls on a horizontal plane, and always remains exactly upright, see for instance [3]. The coordinates (x_1, x_2, θ, ϕ) describe the possible configurations of the system, where (x_1, x_2) are the coordinates of the contact point with the x_1x_2 -plane, θ the heading angle and ϕ the self-rotation angle. The system is shown in Figure 1. Not all the velocities are admissible for this system, since the constraint that the disk roll without slipping is specified by the linear velocity constraints

$$(2) \quad \dot{x}_1 - r\dot{\phi} \cos \theta = 0, \quad \dot{x}_2 - r\dot{\phi} \sin \theta = 0,$$

and therefore,

$$\mathcal{D} = \text{span} \left\{ X_1 = r \cos \theta \frac{\partial}{\partial x_1} + r \sin \theta \frac{\partial}{\partial x_2} + \frac{\partial}{\partial \phi}, X_2 = \frac{\partial}{\partial \theta} \right\}.$$

The constraints are nonholonomic since the distribution \mathcal{D} is not involutive. Indeed,

$$[X_1, X_2] = r \sin \theta \frac{\partial}{\partial x_1} - r \cos \theta \frac{\partial}{\partial x_2},$$

and $[X_1, X_2](q) \notin \mathcal{D}_q$ for all $q \in Q$.

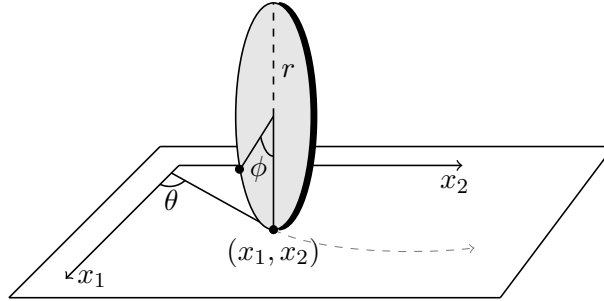


FIGURE 1. The geometry of the rolling disk.

◇

2.1. Lagrangian equations for nonholonomic systems. In addition to the constraints, the dynamics is specified by a Lagrangian function $L : TQ \rightarrow \mathbb{R}$, assumed to be of mechanical type, that is,

$$L(v_q) = \frac{1}{2} g(v_q, v_q) - V(q), \quad v_q \in T_q Q,$$

where g is a Riemannian metric on the configuration space Q and $V : Q \rightarrow \mathbb{R}$ a potential function. The Lagrangian is written in coordinates (q^i, \dot{q}^i) as

$$L(q^i, \dot{q}^i) = \frac{1}{2} g_{ij}(q) \dot{q}^i \dot{q}^j - V(q),$$

where $g_{ij} = g(\partial/\partial q^i, \partial/\partial q^j)$, $1 \leq i, j \leq n$.

Example (Rolling disk, continued). For the case of the rolling disk the Riemannian metric is

$$g = m dx_1 \otimes dx_1 + m dx_2 \otimes dx_2 + J_\theta d\theta \otimes d\theta + J_\phi d\phi \otimes d\phi,$$

where m is the mass of the disk and J_θ and J_ϕ are the moment of inertia about the θ and ϕ axis respectively. We assume that the Lagrangian is purely kinetic, i.e. $V = 0$, and thus

$$L(x_1, x_2, \theta, \phi, \dot{x}_1, \dot{x}_2, \dot{\theta}, \dot{\phi}) = \frac{1}{2} \left(m \dot{x}_1^2 + m \dot{x}_2^2 + J_\theta \dot{\theta}^2 + J_\phi \dot{\phi}^2 \right).$$

◇

In nonholonomic mechanics, the equations of motion are completely determined by the Lagrange-d'Alembert principle. This principle states that a curve $q : I \subset \mathbb{R} \rightarrow Q$ is an admissible motion of the system if

$$\delta \mathcal{J} = \delta \int_0^T L(q(t), \dot{q}(t)) dt = 0,$$

for all variations such that $\delta q(t) \in \mathcal{D}_{q(t)}$, $0 \leq t \leq T$, $\delta q(0) = \delta q(T) = 0$. The velocity of the curve itself must also satisfy the constraints according to (1). From the Lagrange-d'Alembert principle, we arrive at the well-known **nonholonomic equations**

$$(3a) \quad \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) - \frac{\partial L}{\partial q^i} = \lambda_\alpha \mu_i^\alpha,$$

$$(3b) \quad \mu_i^\alpha(q) \dot{q}^i = 0,$$

where λ_α , $m+1 \leq \alpha \leq n$, is a set of Lagrange multipliers. The right-hand side of Equation (3a) represents the force induced by the constraints, while Equation (3b) gives the linear velocity constraints themselves.

It is important to stress that in Equations (3) it is necessary to use the Lagrangian defined on the full space TQ instead of working with the restriction of L to \mathcal{D} (where we now consider \mathcal{D} as a vector subbundle of TQ). Applying standard variational techniques and using $l = L|_{\mathcal{D}}$ we would derive a different set of equations than (3), which are not valid for nonholonomic mechanics. These other equations are called vakonomic equations, or variational constrained equations in the literature, see for instance [1].

2.1.1. Adapted coordinates. Equations (3) are derived using a set of coordinates (q^i) on Q , and the induced coordinates (q^i, \dot{q}^i) on TQ by the canonical coordinate frame $\left\{ \frac{\partial}{\partial q^i} \right\}$, $1 \leq i \leq n$. Any element $v_q \in T_q Q$ can therefore be written univocally as

$$v_q = \dot{q}^i \frac{\partial}{\partial q^i} \Big|_q.$$

In the case of nonholonomic mechanics it can be useful to adapt the chosen frame to the linear velocity constraints. Specifically we consider a basis of vector fields $\{X_a, X_\alpha\}$, $1 \leq a \leq m$ and $m+1 \leq \alpha \leq n$, such that locally

$$\mathcal{D}_q = \text{span}\{X_a(q)\} \quad \text{and} \quad \mathcal{D}_q^{\perp, g} = \text{span}\{X_\alpha(q)\},$$

where $\mathcal{D}_q^{\perp, g}$ is the (Riemannian)-orthogonal to \mathcal{D} , i.e.

$$g(X_a, X_\alpha) = 0, \quad 1 \leq a \leq m \quad \text{and} \quad m+1 \leq \alpha \leq n.$$

Observe that $T_q Q = \mathcal{D}_q \oplus \mathcal{D}_q^{\perp, g}$.

The adapted basis $\{X_a, X_\alpha\}$ induces a new set of coordinates on the tangent bundle (q^i, y^a, y^α) (also called quasi-velocities) so that now

$$v_q = y^a X_a(q) + y^\alpha X_\alpha(q).$$

Observe that the elements $v_q \in \mathcal{D}_q$ are distinguished by $y^\alpha = 0$. Therefore $y^\alpha = 0$ expresses the nonholonomic constraints in the adapted basis. Consequently \mathcal{D} is completely described by coordinates (q^i, y^a) .

Example (Rolling disk, continued). We take an adapted basis $\{X_1, X_2, X_3, X_4\}$, where

$$\begin{aligned}\mathcal{D} &= \text{span} \left\{ X_1 = r \cos \theta \frac{\partial}{\partial x_1} + r \sin \theta \frac{\partial}{\partial x_2} + \frac{\partial}{\partial \phi}, X_2 = \frac{\partial}{\partial \theta} \right\}, \\ \mathcal{D}^{\perp, g} &= \text{span} \left\{ X_3 = \frac{1}{m} \frac{\partial}{\partial x_1} - \frac{r}{J_\phi} \cos \phi \frac{\partial}{\partial \phi}, X_4 = \frac{1}{m} \frac{\partial}{\partial x_2} - \frac{r}{J_\phi} \sin \phi \frac{\partial}{\partial \phi} \right\}.\end{aligned}$$

This induces coordinates $(x_1, x_2, \theta, \phi, y^1, y^2, y^3, y^4)$ on TQ , which are related to the standard coordinates as follows:

$$\begin{aligned}\dot{x}_1 &= r y^1 \cos \theta + \frac{y^3}{m}, \\ \dot{x}_2 &= r y^1 \sin \theta + \frac{y^4}{m}, \\ \dot{\theta} &= y^2, \\ \dot{\phi} &= y^1 - \frac{r}{J_\phi} y^3 \cos \phi - \frac{r}{J_\phi} y^4 \sin \phi.\end{aligned}$$

Observe that the linear constraints have the simple form $y^3 = 0, y^4 = 0$ in the adapted basis. \diamond

2.1.2. Equations of motion in adapted coordinates. We now want to rewrite the equations of motion of the nonholonomic system in terms of the coordinates (q^i, y^a, y^α) , instead of the canonical coordinates (q^i, \dot{q}^i) . Consider first Equation (3a). We can split it as the following system of equations:

$$(4a) \quad 0 = X_a^i \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) - X_a^i \frac{\partial L}{\partial q^i} - \lambda_\beta \mu_i^\beta X_a^i = X_a^i \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) - X_a^i \frac{\partial L}{\partial q^i},$$

$$(4b) \quad 0 = X_\alpha^i \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) - X_\alpha^i \frac{\partial L}{\partial q^i} - \lambda_\beta \mu_i^\beta X_\alpha^i,$$

with $1 \leq a \leq m, m+1 \leq \alpha, \beta \leq n$. In Equation (4a) we have used that $X_a(q) \in \mathcal{D}_q$. Observe that Equation (4b) uniquely gives information about the value of the Lagrange multipliers since g is a Riemannian metric and therefore $(\mu_i^\beta X_\alpha^i)$ is a regular matrix. As we are not interested in the evolution of the Lagrange multipliers we discard this second set of equations.

Define the Lagrangian in adapted coordinates as $\tilde{L}(q^i, y^a, y^\alpha) := L(q^i, X_a^i y^a + X_\alpha^i y^\alpha)$. We want to express Equation (4a) in terms of \tilde{L} . To this end observe that

$$\begin{aligned}\frac{\partial \tilde{L}}{\partial y^a} &= X_a^i \frac{\partial L}{\partial \dot{q}^i}, \\ \frac{\partial \tilde{L}}{\partial y^\alpha} &= X_\alpha^i \frac{\partial L}{\partial \dot{q}^i}, \\ \frac{\partial \tilde{L}}{\partial q^j} &= \frac{\partial L}{\partial q^j} + \left(y^a \frac{\partial X_a^i}{\partial q^j} + y^\alpha \frac{\partial X_\alpha^i}{\partial q^j} \right) \frac{\partial L}{\partial \dot{q}^i}.\end{aligned}$$

Now define the restricted Lagrangian $l : \mathcal{D} \rightarrow \mathbb{R}$ by $l := \tilde{L}|_{\mathcal{D}}$, that is, $l(q^i, y^a) := \tilde{L}(q^i, y^a, 0)$. It is interesting to note that

$$\tilde{L}(q^i, y^a, y^\alpha) = \frac{1}{2} g_{ab} y^a y^b + \frac{1}{2} g_{\alpha\beta} y^\alpha y^\beta - V(q),$$

where $g_{ab} := g(X_a, X_b)$ and $g_{\alpha\beta} := g(X_\alpha, X_\beta)$, and thus

$$l(q^i, y^a) = \frac{1}{2} g_{ab} y^a y^b - V(q).$$

We will make use of the fact that we can express the bracket $[X_a, X_b]$ in two ways using the different frames, concretely as

$$\begin{aligned} [X_a, X_b] &= \left(\frac{\partial X_b^j}{\partial q^i} X_a^i - \frac{\partial X_a^j}{\partial q^i} X_b^i \right) \frac{\partial}{\partial q^j} = [X_a, X_b]^j \frac{\partial}{\partial q^j}, \\ [X_a, X_b] &= C_{ab}^c X_c + C_{ab}^\alpha X_\alpha. \end{aligned}$$

Now, taking the restriction of Equation (4a) to \mathcal{D} , that is, using that $y^\alpha = 0$, we get

$$\begin{aligned} 0 &= X_a^i \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) - X_a^i \frac{\partial L}{\partial q^i} \\ &= \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} X_a^i \right) - \frac{dX_a^i}{dt} \frac{\partial L}{\partial \dot{q}^i} - X_a^i \frac{\partial L}{\partial q^i} \\ &= \frac{d}{dt} \left(\frac{\partial l}{\partial y^a} \right) + [X_a, X_b]^i y^b \frac{\partial L}{\partial \dot{q}^i} - X_a^i \frac{\partial l}{\partial q^i}. \end{aligned}$$

The middle term of the last equation is

$$\begin{aligned} [X_a, X_b]^i y^b \frac{\partial L}{\partial \dot{q}^i} &= (C_{ab}^c X_c^i + C_{ab}^\alpha X_\alpha^i) y^b \frac{\partial L}{\partial \dot{q}^i} \\ &= C_{ab}^c y^b \frac{\partial l}{\partial y^c} + C_{ab}^\alpha y^b \frac{\partial \tilde{L}}{\partial y^\alpha} \\ &= C_{ab}^c y^b \frac{\partial l}{\partial y^c}, \end{aligned}$$

since $\partial \tilde{L} / \partial y^\alpha = g_{\alpha\beta} y^\beta = 0$ because $y^\alpha = 0$. In conclusion we have that

$$\begin{aligned} 0 &= X_a^i \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) - X_a^i \frac{\partial L}{\partial q^i} \\ &= \frac{d}{dt} \left(\frac{\partial l}{\partial y^a} \right) + C_{ab}^c y^b \frac{\partial l}{\partial y^c} - X_a^i \frac{\partial l}{\partial q^i}. \end{aligned}$$

Therefore, the equations of motion of the nonholonomic system are rewritten in terms of the restricted Lagrangian l as

$$(5a) \quad \frac{d}{dt} \left(\frac{\partial l}{\partial y^a} \right) + C_{ab}^c y^b \frac{\partial l}{\partial y^c} - X_a^i \frac{\partial l}{\partial q^i} = 0,$$

$$(5b) \quad \dot{q}^i = X_a^i(q) y^a,$$

see for instance [18, 29].

Example (Rolling disk, continued). We have the restricted Lagrangian

$$l(x_1, x_2, \theta, \phi, y^1, y^2) = \frac{1}{2} [(mr^2 + J_\phi)(y^1)^2 + J_\theta(y^2)^2].$$

Now observe that

$$\begin{aligned} [X_1, X_2] &= r \sin \theta \frac{\partial}{\partial x_1} - r \cos \theta \frac{\partial}{\partial x_2} \\ &= mr \sin \theta X_3 - mr \cos \theta X_4. \end{aligned}$$

Therefore in this simple example we have $C_{ab}^c = 0$ for all $1 \leq a, b, c \leq 2$. The equations of motion (5) for this nonholonomic system are

$$\begin{aligned} \dot{x}_1 &= ry^1 \cos \theta, & \dot{\theta} &= y^2, & \dot{y}^1 &= 0, \\ \dot{x}_2 &= ry^1 \sin \theta, & \dot{\phi} &= y^1, & \dot{y}^2 &= 0, \end{aligned}$$

which are immediately explicitly integrated. ◇

2.2. “Hamiltonian equations” for nonholonomic systems. On the cotangent bundle T^*Q the Lagrangian is replaced by the corresponding Hamiltonian H . We still assume a mechanical system, and let (q^i, p_i) , $1 \leq i \leq n$, give local canonical coordinates on T^*Q through the Legendre transformation $\mathcal{FL} : TQ \rightarrow T^*Q$, i.e.

$$\mathcal{FL} : (q^i, \dot{q}^i) \mapsto (q^i, p_i = \partial L / \partial \dot{q}^i).$$

Then H is given locally by

$$H(q^i, p_i) = \frac{1}{2} p_i g^{ij} p_j + V(q),$$

with (g^{ij}) being the inverse matrix of (g_{ij}) .

The Hamiltonian form of the nonholonomic equations (3) is then

$$(6a) \quad \begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = J \begin{pmatrix} \frac{\partial H}{\partial q}(q, p) \\ \frac{\partial H}{\partial p}(q, p) \end{pmatrix} + \lambda_\alpha \begin{pmatrix} 0 \\ \mu^\alpha(q) \end{pmatrix},$$

$$(6b) \quad \mu_i^\alpha(q) \frac{\partial H}{\partial p_i}(q, p) = \mu_i^\alpha g^{ik} p_k = 0,$$

where $m+1 \leq \alpha \leq n$ and $J = \begin{pmatrix} 0_n & I_n \\ -I_n & 0_n \end{pmatrix}$, see e.g. [34].

2.2.1. Equations of motion in adapted coordinates. Now we will rewrite the restricted nonholonomic equations in a Hamilton-like way on D^* (see [2, 10, 18]). More precisely, consider the Legendre transformation $\mathcal{Fl} : \mathcal{D} \rightarrow \mathcal{D}^*$, locally given by

$$\mathcal{Fl} : (q^i, y^a) \mapsto (q^i, \rho_a = \frac{\partial l}{\partial y^a}),$$

From the Legendre transformation we can define the Hamiltonian function $\mathcal{H} : \mathcal{D}^* \rightarrow \mathbb{R}$ which in local coordinates becomes

$$\mathcal{H}(q^i, \rho_a) = \frac{1}{2} g^{ab} \rho_a \rho_b + V(q).$$

Then upon changing coordinates in (5) using the Legendre transform and \mathcal{H} , the equations of motion of a nonholonomic system are equivalently rewritten as

$$(7a) \quad \dot{q}^i = X_b^i \frac{\partial \mathcal{H}}{\partial \rho_b},$$

$$(7b) \quad \dot{\rho}_a = -C_{ab}^c \rho_c \frac{\partial \mathcal{H}}{\partial \rho_b} - X_a^i \frac{\partial \mathcal{H}}{\partial q^i}.$$

If we define the skew-symmetric matrix

$$(8) \quad \Pi(q, \rho) = \begin{pmatrix} 0 & X_b^i \\ -(X_a^j)^T & -C_{ab}^c \rho_c \end{pmatrix}$$

then the equations (7) will be given by

$$(9) \quad \dot{\zeta} = \Pi(\zeta) \nabla \mathcal{H}(\zeta),$$

where $\zeta = (q^i, \rho_a)$ are coordinates on \mathcal{D}^* . This skew gradient format will allow the use of discrete gradient methods, as we will see in the next section.

Remark 2.1. It is possible to give a more intrinsic definition of these objects, as in [10]. Denote by $\{\cdot, \cdot\}$ the canonical bracket of the cotangent bundle T^*Q . Define a bracket of functions $\{\cdot, \cdot\}_{\mathcal{D}^*}$ on D^* by

$$\{f, g\}_{\mathcal{D}^*} = \{f \circ i_{\mathcal{D}}^*, g \circ i_{\mathcal{D}}^*\} \circ P^*,$$

for $f, g \in C^\infty(\mathcal{D}^*)$ where $i_{\mathcal{D}}^* : T^*Q \rightarrow \mathcal{D}^*$ and $P^* : \mathcal{D}^* \rightarrow T^*Q$ are the dual maps of the monomorphisms $i_{\mathcal{D}} : \mathcal{D} \rightarrow TQ$ and the projector $P : TQ \rightarrow \mathcal{D}$, respectively. Then the bivector field Π is given by

$$\Pi(df, dg) = \{f, g\}_{\mathcal{D}^*}.$$

This bracket does not in general satisfy the Jacobi identity, that is

$$\{f, \{g, h\}_{\mathcal{D}^*}\}_{\mathcal{D}^*} + \{g, \{h, f\}_{\mathcal{D}^*}\}_{\mathcal{D}^*} + \{h, \{f, g\}_{\mathcal{D}^*}\}_{\mathcal{D}^*} \neq 0.$$

By using this bracket, Equation (9) will be more appropriately written as

$$\dot{f} = \{f, \mathcal{H}\}_{\mathcal{D}^*} \text{ for all } f \in C^\infty(\mathcal{D}^*).$$

Example (Rolling disk, continued). We have the Hamiltonian function

$$\mathcal{H}(x_1, x_2, \theta, \phi, \rho_1, \rho_2) = \frac{1}{2} \left[\frac{\rho_1^2}{mr^2 + J_\phi} + \frac{\rho_2^2}{J_\theta} \right],$$

where $\rho_1 = (mr^2 + J_\phi)y^1$ and $\rho_2 = J_\theta y^2$. The skew-symmetric matrix (almost-Poisson structure) (8) is given by

$$\Pi = \begin{pmatrix} 0 & 0 & 0 & 0 & r \cos \theta & 0 \\ 0 & 0 & 0 & 0 & r \sin \theta & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ -r \cos \theta & -r \sin \theta & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \end{pmatrix},$$

and the equations of motion (9) are

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{\theta} \\ \dot{\phi} \\ \dot{\rho}_1 \\ \dot{\rho}_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & r \cos \theta & 0 \\ 0 & 0 & 0 & 0 & r \sin \theta & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ -r \cos \theta & -r \sin \theta & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{\rho_1}{mr^2 + J_\phi} \\ \frac{\rho_2}{J_\theta} \end{pmatrix}.$$

◇

3. ENERGY-PRESERVING INTEGRATORS BASED ON DISCRETE GRADIENTS

In the previous section, we reduced the study of the nonholonomic dynamics to a system of differential equations

$$\dot{\zeta} = \Pi(\zeta) \nabla \mathcal{H}(\zeta)$$

on \mathcal{D}^* . In this section we will assume that Q is a real vector space, therefore $\mathcal{D}^* \cong \mathbb{R}^N$ where $n + m = N$. For a generalisation to the case of Lie groups and homogeneous manifolds see [5].

Since nonholonomic dynamics does not preserve the almost-Poisson structure Π in general, we will focus on the preservation of the energy using geometric integrators which preserve exactly this quantity. In particular, we will use discrete analogues of the gradient of the Hamiltonian function [26].

3.1. Discrete gradients. For ODEs in skew-gradient form, i.e. $\dot{x} = \Pi(x) \nabla H(x)$ with $x \in \mathbb{R}^N$ and $\Pi(x)$ a skew-symmetric matrix, it is immediate to check that H is a first integral. Indeed

$$\dot{H} = \nabla H(x)^T \dot{x} = \nabla H(x)^T \Pi(x) \nabla H(x) = 0,$$

due to the skew-symmetry of Π . Using discretizations of the gradient $\nabla H(x)$ it is possible to define a class of integrators which preserve the first integral H exactly.

Definition 3.1. [17] *Let $H : \mathbb{R}^N \rightarrow \mathbb{R}$ be a differentiable function. Then $\bar{\nabla} H : \mathbb{R}^{2N} \rightarrow \mathbb{R}^N$ is a discrete gradient of H if it is continuous and satisfies*

$$(10a) \quad \bar{\nabla} H(x, x')^T (x' - x) = H(x') - H(x), \quad \text{for all } x, x' \in \mathbb{R}^N,$$

$$(10b) \quad \bar{\nabla} H(x, x) = \nabla H(x), \quad \text{for all } x \in \mathbb{R}^N.$$

Some well-known examples of discrete gradients are:

- The mean value (or averaged vector field) discrete gradient introduced in [20] and given by

$$(11) \quad \bar{\nabla}_1 H(x, x') := \int_0^1 \nabla H((1 - \xi)x + \xi x') d\xi, \quad \text{for } x' \neq x.$$

- The midpoint (or Gonzalez) discrete gradient, introduced in [17] and given by

$$(12) \quad \bar{\nabla}_2 H(x, x') := \nabla H\left(\frac{1}{2}(x' + x)\right) + \frac{H(x') - H(x) - \nabla H\left(\frac{1}{2}(x' + x)\right)^T (x' - x)}{|x' - x|^2} (x' - x),$$

for $x' \neq x$.

- The coordinate increment discrete gradient, introduced in [21], with each component given by

$$(13) \quad \bar{\nabla}_3 H(x, x')_i = \frac{H(x'_1, \dots, x'_i, x_{i+1}, \dots, x_n) - H(x'_1, \dots, x'_{i-1}, x_i, \dots, x_n)}{x'_i - x_i}, \quad 1 \leq i \leq N,$$

when $x'_i \neq x_i$, and $\bar{\nabla}_3 H(x, x')_i = \frac{\partial H}{\partial x_i}(x'_1, \dots, x'_{i-1}, x'_i = x_i, x_{i+1}, \dots, x_n)$ otherwise.

It can be easily checked that these are indeed discrete gradients, see [17], [21] and [26].

3.2. Integrators based on discrete gradients. Once a discrete gradient $\bar{\nabla}H$ has been chosen, it is straightforward to define an energy-preserving integrator by

$$(14) \quad \frac{x' - x}{h} = \tilde{\Pi}(x, x', h) \bar{\nabla}H(x, x'),$$

where $\tilde{\Pi}$ is a differentiable skew-symmetric matrix approximating Π , that is, it satisfies $\tilde{\Pi}(x, x, 0) = \Pi(x)$. As in the continuous case, it is immediate to check that H is exactly preserved, since

$$H(x') - H(x) = \bar{\nabla}H(x, x')^T (x' - x) = h \bar{\nabla}H(x, x')^T \tilde{\Pi}(x, x', h) \bar{\nabla}H(x, x') = 0.$$

If we further wish to get a second order method then it is sufficient to choose $\tilde{\Pi}$ such that $\tilde{\Pi}(x, x', h) = \tilde{\Pi}(x', x, -h)$, and a differentiable discrete gradient such that $\bar{\nabla}H(x, x') = \bar{\nabla}H(x', x)$. This guarantees that the integration method (14) is time-symmetric and therefore second order accurate, see [19]. For instance it is enough to choose $\tilde{\Pi}(x, x', h) = \Pi\left(\frac{x+x'}{2}\right)$ and take the mean value discrete gradient or the midpoint discrete gradient. Higher order energy-preserving methods, which generalize the mean value discrete gradient (11), can be obtained by collocation methods as in [6].

Remark 3.2. If the Hamiltonian is quadratic then

$$\bar{\nabla}_1 H(x, x') = \bar{\nabla}_2 H(x, x') = \nabla H\left(\frac{1}{2}(x' + x)\right),$$

that is, the mean value discrete gradient (11) and the Gonzalez discrete gradient (12) coincide with the continuous gradient evaluated at the midpoint. Then if we choose $\tilde{\Pi}(x, x', h) = \Pi\left(\frac{x+x'}{2}\right)$ the method (14) reduces to the implicit midpoint rule. If the Hamiltonian is of the form $H(x) = \sum_{j=1}^N a_j x_j^2$, then $(\nabla H)_j = 2a_j x_j$ and

$$\bar{\nabla}_1 H(x, x')_i = \bar{\nabla}_2 H(x, x')_i = \bar{\nabla}_3 H(x, x')_i = a_i(x'_i + x_i), \quad 1 \leq i \leq N,$$

that is, all three discrete gradients introduced above coincide with $\nabla H\left(\frac{1}{2}(x' + x)\right)$.

Remark 3.3. Preservation of the nonholonomic constraints. Going back to the case of nonholonomic systems, we can now apply an energy-preserving method (14) to Equation (9). Notice that if we take the approximation $\tilde{\Pi}(\zeta, \zeta', h)$ to be $\Pi(\bar{\zeta})$ for some $\bar{\zeta}(\zeta, \zeta') \in D^*$, and let $\partial \bar{\mathcal{H}} / \partial \rho_a$ be the discrete gradient component that approximates $\partial \mathcal{H} / \partial \rho_a$, then a discrete gradient method (14) gives

$$\frac{q'^j - q^j}{h} = X_a^j(\bar{q}) \frac{\partial \bar{\mathcal{H}}}{\partial \rho_a}(\zeta, \zeta'),$$

where $\bar{q} = \pi_Q(\bar{\zeta})$. When applying $\mu^\alpha \in D^\circ$ we obtain

$$\mu_j^\alpha \left(\frac{q'^j - q^j}{h} \right) = \mu_j^\alpha X_a^j(\bar{q}) \frac{\partial \bar{\mathcal{H}}}{\partial \rho_a}(\zeta, \zeta') = 0,$$

since $\mu_j^\alpha X_a^j = 0$ for all $1 \leq a \leq m$, $m+1 \leq \alpha \leq n$. All the nonholonomic constraints are thus preserved by the method.

Example (Rolling disk, continued). Using any of the three discrete gradients introduced in Section 3.1 and a midpoint approximation of Π , we get the following energy preserving integrator, which is precisely the implicit midpoint rule:

$$\begin{aligned} x'_1 &= x_1 + hr \cos\left(\frac{\theta + \theta'}{2}\right) \frac{\rho_1}{mr^2 + J_\phi}, & \theta' &= \theta + h \frac{\rho_2}{J_\theta}, & \rho'_1 &= \rho_1, \\ x'_2 &= x_2 + hr \sin\left(\frac{\theta + \theta'}{2}\right) \frac{\rho_1}{mr^2 + J_\phi}, & \phi' &= \phi + h \frac{\rho_1}{mr^2 + J_\phi}, & \rho'_2 &= \rho_2. \end{aligned}$$

Observe that, as a consequence, we deduce the preservation of the constraints

$$\frac{x'_1 - x_1}{h} - r \left(\frac{\phi' - \phi}{h} \right) \cos\left(\frac{\theta + \theta'}{2}\right) = 0, \quad \frac{x'_2 - x_2}{h} - r \left(\frac{\phi' - \phi}{h} \right) \sin\left(\frac{\theta + \theta'}{2}\right) = 0,$$

which are discretizations of the nonholonomic constraints (2). \diamond

3.3. Integrators on T^*Q . The equations of motion in adapted coordinates for a given nonholonomic system are usually not known initially. A potential obstacle in applying a discrete gradient method to the equations in adapted coordinates (9) is therefore that the user must analytically derive these equations. In this section we formulate the proposed schemes directly on the Hamiltonian equations of motion in canonical coordinates (6), and achieve preservation of energy and the nonholonomic constraints without explicitly deriving and using the formulation in adapted coordinates.

As a first attempt at an energy preserving method, we can define a numerical integrator for (6) directly on T^*Q by

$$(15) \quad \frac{z' - z}{h} = J \bar{\nabla} H(z, z') + \lambda_\alpha \begin{pmatrix} 0 \\ \mu^\alpha(\bar{q}) \end{pmatrix}, \quad \bar{\nabla} H(z, z')^T \begin{pmatrix} 0 \\ \mu^\alpha(\bar{q}) \end{pmatrix} = 0,$$

where $\bar{\nabla} H$ is a discrete gradient, $z = (q, p)$ and $z' = (q', p')$. Notice that this method is energy-preserving, since

$$\begin{aligned} H(z') - H(z) &= \bar{\nabla} H(z, z')^T (z' - z) \\ &= h \bar{\nabla} H(z, z')^T J \bar{\nabla} H(z, z') + h \lambda_\alpha \bar{\nabla} H(z, z')^T \begin{pmatrix} 0 \\ \mu^\alpha(\bar{q}) \end{pmatrix} = 0. \end{aligned}$$

However the constraints (6b) will in general only be approximately satisfied at the solution points by such a method.

To achieve exact preservation of both the energy and the nonholonomic constraints (6b), we utilize the restricted equations (9) on \mathcal{D}^* .

The method to step from (q, p) to (q', p') can be summarized as:

- (i) Change coordinates from (q, p) to (q, ρ) .
- (ii) Step from (q, ρ) to (q', ρ') using a discrete gradient method (14) applied to (9).
- (iii) Change coordinates from (q', ρ') to (q', p') .

For step (i) and (iii) we make use of the following relations between the coordinates ρ_b on \mathcal{D}^* and p_i on $\mathcal{FL}(\mathcal{D})$

$$(16a) \quad p_i = \mathbf{g}_{ij} X_a^j g^{ab} \rho_b, \quad 1 \leq i \leq n,$$

$$(16b) \quad \rho_b = X_b^i p_i, \quad 1 \leq b \leq m.$$

The challenge is performing step (ii) without any explicit knowledge of the equations in adapted coordinates. Specifically we need to evaluate $\nabla \mathcal{H}$ and Π in (9) for any $\zeta \in \mathcal{D}^*$. Let us therefore rewrite these expressions in a suitable format. Here p is considered a dependent variable of (q, ρ) through (16a).

First observe that the skew symmetric matrix $-C_{ab}^c \rho_c$ in (8) may be written as

$$(17) \quad -C_{ab}^c \rho_c = -C_{ab}^c X_c^j p_j = -[X_a, X_b]^j p_j = \left(\frac{\partial X_a^j}{\partial q^i} X_b^i - \frac{\partial X_b^j}{\partial q^i} X_a^i \right) p_j.$$

Second we can write the partial derivatives of \mathcal{H} as follows

$$(18a) \quad \begin{aligned} \frac{\partial \mathcal{H}}{\partial q^i} &= \frac{1}{2} p_j \frac{\partial \mathbf{g}^{jk}}{\partial q^i} p_k + \frac{\partial p_j}{\partial q^i} \mathbf{g}^{jk} p_k + \frac{\partial V}{\partial q^i} = \frac{1}{2} p_j \frac{\partial \mathbf{g}^{jk}}{\partial q^i} p_k + \frac{\partial p_j}{\partial q^i} X_a^j g^{ab} \rho_b + \frac{\partial V}{\partial q^i} \\ &= \frac{1}{2} p_j \frac{\partial \mathbf{g}^{jk}}{\partial q^i} p_k - p_j \frac{\partial X_a^j}{\partial q^i} g^{ab} \rho_b + \frac{\partial V}{\partial q^i}, \end{aligned}$$

$$(18b) \quad \frac{\partial \mathcal{H}}{\partial \rho_a} = g^{ab} \rho_b = g^{ab} X_b^i p_i.$$

Expressing (9) using (17) and (18), the remaining issue is that we don't have explicit knowledge of a basis $X_a(q)$ for the distribution \mathcal{D} or of the partial derivatives $\partial X_a(q)/\partial q^i$. For an arbitrary point q we generate $X_a(q)$ by computing the QR-factorization of the constraint matrix $(\mu_i^\alpha(q))$ using Householder reflections, see e.g. [16]. The last m columns of the Q matrix can then be taken as $X_a(q)$, $1 \leq a \leq m$. Householder reflections were chosen because they are numerically stable and efficient.

We now make the assumption that the partial derivatives of $(\mu_i^\alpha(q))$ are either known or easily derived, which is usually the case. Then $\partial X_a(q)/\partial q^i$ can be calculated by augmenting the QR-factorization algorithm with corresponding steps for the partial derivatives. The procedure is specified in Algorithm 1 for the matrix $A(q) := (\mu_i^\alpha(q))$.

To ensure we are sampling the same basis vector fields at different points in a given step when using Algorithm 1, it is sufficient to make sure the vector of sign choices s remains fixed for all factorizations in a given integration step. Because we only suppose knowledge of the full system (6), we transform back to canonical coordinates (q, p) after each step.

In theory this implementation can be combined with any discrete gradient method. However, since it is desirable to minimize the number of QR-factorizations per time step, this approach is best suited when used together with the Gonzalez discrete gradient and a midpoint approximation of Π . We shall refer to this specific method later as GONZALEZ-R.

Remark 3.4. Computational cost. For the initial direct method (15) it is necessary to evaluate the Lagrange multipliers. Moreover it is necessary to implement the constraint equations in each step of the algorithm. Applying a discrete gradient method (14) directly to the reduced system (9) simplifies the computational cost. This is so because the constraints are preserved automatically, and it is not necessary to compute the Lagrange multipliers as additional variables. Specifically with the method (15) it is necessary to solve $3n - m$ variables while using (14) on (9) it is only necessary to compute $n + m$ variables.

Integrating the full system using the equations in adapted coordinates and the QR-factorization approach, we avoid the problem with Lagrange multipliers, but still see a rise in computational cost due to the necessity of moving between coordinate systems, and the general added cost in evaluating $\nabla \mathcal{H}$ and Π .

Thus the trade-off in not requiring knowledge of the reduced system is an increase in computational cost. It is therefore generally more efficient to analytically derive (9) and apply a discrete gradient method (14).

Remark 3.5. Implementation using finite differences. It is remarked in [24] that the condition (10b) for discrete gradients is only required to ensure consistency. Suppose this condition is relaxed slightly to

$$\bar{\nabla} H(x, x, h) = \nabla H(x) + \mathcal{O}(h^r),$$

where r should at least match the order of the method. This is sufficient for the consistency of an integrator (14), and indeed for the method to have order r . We can use this to avoid having to evaluate $\partial X_a(q)/\partial q^i$ at the midpoint in GONZALEZ-R by replacing it with an appropriate finite difference approximation,

Algorithm 1 (QR factorization procedure with differentiation using Householder reflections). Computes the QR factorization of a differentiable matrix $A(q) \in \mathbb{R}^{n,n-m}$ as well as all partial derivatives $\partial Q/\partial q^i$, $1 \leq i \leq n$, for any $q \in \mathbb{R}^n$. Let $\partial_i := \partial/\partial q^i$ while $B_{i:j,k:l}$ with $i \leq j$ and $k \leq l$ denotes the submatrix containing rows i to j and columns k to l of a matrix B , with the shorthand $i := i : i$. ∂B denotes the tensor containing all partial derivatives of B at q . $s \in \{+1, -1\}^{n-m}$ is a vector of sign choices.

```

procedure QRDIFF( $A, \partial A, n, m, s$ )
   $Q^{(0)} \leftarrow I_n$ 
   $R^{(0)} \leftarrow A$ 
  for  $i = 1, 2, \dots, n$  do
     $\partial_i Q^{(0)} \leftarrow 0_n$ 
     $\partial_i R^{(0)} \leftarrow \partial_i A$ 
  end for

  for  $k = 0, 1, \dots, n-m-1$  do
    for  $j = 1, 2, \dots, n$  do
      if  $j \leq k$  then
         $\tilde{w}_j^{(k)} \leftarrow 0$ 
      else if  $j = k+1$  then
         $\tilde{w}_j^{(k)} \leftarrow R_{jk}^{(k)} + s_k \|R_{k:n,k}^{(k)}\|_2$ 
      else
         $\tilde{w}_j^{(k)} \leftarrow R_{jk}^{(k)}$ 
      end if
    end for
     $\|\tilde{w}^{(k)}\|_2 \leftarrow \sqrt{2 \left( \|R_{k:n,k}^{(k)}\|_2^2 + s_k R_{kk}^{(k)} \|R_{k:n,k}^{(k)}\|_2 \right)}$ 
     $w^{(k)} \leftarrow \frac{\tilde{w}^{(k)}}{\|\tilde{w}^{(k)}\|_2}$ 
     $u^{(k)} \leftarrow w^{(k)T} R^{(k)}$ 
     $R^{(k+1)} \leftarrow R^{(k)} - 2w^{(k)}u^{(k)}$ 
     $Q^{(k+1)} \leftarrow Q^{(k)} - 2Q^{(k)} \left( w^{(k)}w^{(k)T} \right)$ 

    for  $i = 1, 2, \dots, n$  do
       $\partial_i \|R_{k:n,k}^{(k)}\|_2 \leftarrow \frac{(\partial_i R_{k:n,k}^{(k)})^T R_{k:n,k}^{(k)}}{\|R_{k:n,k}^{(k)}\|_2}$ 
      for  $j = 1, 2, \dots, n$  do
        if  $j \leq k$  then
           $\partial_i \tilde{w}_j^{(k)} \leftarrow 0$ 
        else if  $j = k+1$  then
           $\partial_i \tilde{w}_j^{(k)} \leftarrow \partial_i R_{jk}^{(k)} + s_k \partial_i \|R_{k:n,k}^{(k)}\|_2$ 
        else
           $\partial_i \tilde{w}_j^{(k)} \leftarrow \partial_i R_{jk}^{(k)}$ 
        end if
      end for
       $\partial_i \|\tilde{w}^{(k)}\|_2 \leftarrow \frac{(\partial_i \tilde{w}^{(k)})^T \tilde{w}^{(k)}}{\|\tilde{w}^{(k)}\|_2}$ 
       $\partial_i w^{(k)} \leftarrow \frac{\partial_i \tilde{w}^{(k)} \|\tilde{w}^{(k)}\|_2 - \tilde{w}^{(k)} \partial_i \|\tilde{w}^{(k)}\|_2}{\|\tilde{w}^{(k)}\|_2^2}$ 
       $\partial_i u^{(k)} \leftarrow \partial_i w^{(k)T} R^{(k)} + w^{(k)T} \partial_i R^{(k)}$ 
       $\partial_i R^{(k+1)} \leftarrow \partial_i R^{(k)} -$ 
         $2 \left( \partial_i w^{(k)} u^{(k)} + w^{(k)} \partial_i u^{(k)} \right)$ 
       $\partial_i Q^{(k+1)} \leftarrow \partial_i Q^{(k)} - 2 \partial_i Q^{(k)} \left( w^{(k)} w^{(k)T} \right) +$ 
         $2 Q^{(k)} \left( \partial_i w^{(k)} w^{(k)T} + w^{(k)} \partial_i w^{(k)T} \right)$ 
    end for
  end for
  return  $Q^{(n-m)}, \partial Q^{(n-m)}$ 
end procedure

```

e.g. the central difference approximation

$$\frac{\partial \bar{X}_a^j}{\partial q^i}(q) := \frac{X_a^j(q + he_i) - X_a^j(q - he_i)}{2h} = \frac{\partial X_a^j}{\partial q^i}(q) + \mathcal{O}(h^2),$$

where e_i is the canonical unit vector i . The resulting method retains second order and still preserves energy and the nonholonomic constraints.

4. EXAMPLES AND NUMERICAL EXPERIMENTS

In this section we apply discrete gradient methods to some illustrative examples of nonholonomic systems. In the first three examples we will derive Equations (9) analytically. In the last one we will compare the strategies proposed in Section 3.3.

4.1. A fully chaotic nonholonomic system. In [24] the authors remark that the key geometric properties for nonholonomic dynamics are not known for general nonintegrable systems. To compare integration methods for such systems, they focus on energy preservation, looking at the following chaotic quartic mechanical system on $Q = \mathbb{R}^{2n+1}$ with coordinates $q = (q^1, q^2, \dots, q^{2n+1})^T := (x, w_1, \dots, w_n, z_1, \dots, z_n)^T$, which is defined by the Lagrangian

$$(19) \quad L(q, \dot{q}) = K(q, \dot{q}) - V(q), \quad \text{where} \quad K = \frac{1}{2} \|\dot{q}\|_2^2, \quad V = \frac{1}{2} \left(\|q\|_2^2 + z_1^2 z_2^2 + \sum_{i=1}^n w_i^2 z_i^2 \right),$$

with the single nonholonomic linear velocity constraint

$$(20) \quad \dot{x} + \sum_{i=1}^n w_i \dot{z}_i = 0.$$

This system is reversible and preserves energy, i.e. $\dot{H} = 0$.

4.1.1. *Lagrangian side.* To derive the equations in adapted coordinates, first note that this Lagrangian is of mechanical type. We can write the kinetic energy K as

$$K(q, \dot{q}) = \frac{1}{2} g_q(\dot{q}, \dot{q}),$$

with the canonical Riemannian metric

$$g = \sum_{i=1}^{2n+1} dq^i \otimes dq^i,$$

which does not depend on q .

The distribution \mathcal{D} and its orthogonal complement $\mathcal{D}^{\perp, g}$ are given by, respectively, the span of $2n$ and 1 independent vector fields:

$$\begin{aligned} \mathcal{D} &= \text{span} \left\{ X_i := \frac{\partial}{\partial w_i}, X_{n+i} := w_i \frac{\partial}{\partial x} - \frac{\partial}{\partial z_i}, i = 1, \dots, n \right\}, \\ \mathcal{D}^{\perp, g} &= \text{span} \left\{ X_{2n+1} := \frac{\partial}{\partial x} + \sum_{i=1}^n w_i \frac{\partial}{\partial z_i} \right\}. \end{aligned}$$

The adapted basis $\{X_1, X_2, \dots, X_{2n+1}\}$, induces new coordinates (q^i, y^a, y^{2n+1}) , $1 \leq i \leq 2n+1$, $1 \leq a \leq 2n$ on TQ for which the nonholonomic constraint reduces to $y^{2n+1} = 0$. The restricted Lagrangian $l : \mathcal{D} \rightarrow \mathbb{R}$ for this system is

$$l(q^i, y^a) = \frac{1}{2} \left(g_{ab} y^a y^b - V(q) \right),$$

where

$$(g_{ab}) = \begin{pmatrix} I_n & 0_n \\ 0_n & I_n + ww^T \end{pmatrix} \quad \text{and} \quad w = (w_1, w_2, \dots, w_n)^T.$$

4.1.2. *Hamiltonian side.* Moving to the Hamiltonian side we replace the velocities in adapted coordinates y^a with the momenta

$$\rho_a = \frac{\partial l}{\partial y^a} = g_{ab} y^b, \quad a = 1, \dots, 2n.$$

The restricted Hamiltonian $\mathcal{H} : \mathcal{D}^* \rightarrow \mathbb{R}$ is then

$$\mathcal{H}(q^i, \rho_a) = \frac{1}{2} g^{ab} \rho_a \rho_b + V(q),$$

where (g^{ab}) is the inverse of (g_{ab}) , i.e.

$$(g^{ab}) = \begin{pmatrix} I_n & 0_n \\ 0_n & I_n - \frac{ww^T}{1+w^T w} \end{pmatrix}.$$

The equations of motion on the Hamiltonian side are then given by (9) where

$$(21) \quad \Pi(q, \rho) = \begin{pmatrix} & 0_{2n+1} & 0_{1 \times n} & w^T \\ & & I_n & 0_n \\ & & 0_n & -I_n \\ 0_{n \times 1} & -I_n & 0_n & 0_n & -\kappa I_n \\ -w & 0_n & I_n & \kappa I_n & 0_n \end{pmatrix}, \quad \nabla \mathcal{H}(q, \rho) = \begin{pmatrix} x \\ w_1 + w_1 z_1^2 - \kappa \rho_{n+1} + \kappa^2 w_1 \\ \vdots \\ w_n + w_n z_n^2 - \kappa \rho_{n+n} + \kappa^2 w_n \\ z_1 + y_1^2 z_1 + z_1 z_2^2 \\ z_2 + y_2^2 z_2 + z_1^2 z_2 \\ z_3 + y_3^2 z_3 \\ \vdots \\ z_n + y_n^2 z_n \\ g^{ab} \rho_b \end{pmatrix},$$

$$\kappa := \frac{w^T \eta}{1 + w^T w} \text{ and } \eta := (\rho_{n+1}, \dots, \rho_{2n})^T.$$

4.1.3. Numerical experiments. We follow the approach in [24], and integrate this system, with $n = 3$, from a random initial state with energy $H = 3.06$. We here compare five different methods. The first two are variational integrators based on the discrete Lagrange d'Alembert (DLA) for the full system (19) and (20). The semi implicit reversible DLA variational integrator proposed in [24] (SI-DLA), and the implicit reversible DLA variational integrator based on a midpoint discrete Lagrangian (I-DLA) which is also described in [24] among others. The third method is the 2-stage Lobatto IIIA-B-C-C*-D SPARK method described in [22] for index 2 DAEs (SPARK), which again discretize the equations of motion of the full system. For the last two methods, we integrate the reduced system, (9) with (21), using a discrete gradient method (14), with two different discrete gradients: The averaged vector field discrete gradient (AVF) (11), and the Gonzalez discrete gradient (GONZALEZ) (12).

As seen in Figure 2, while all five methods are known to be second order accurate and respect the constraint, only the discrete gradient methods conserve the energy up to round off error. In [24] it is shown that the energy error for SI-DLA closely follows a random walk with the variance $\sigma^2 = 10^{-4} h^4 t$. In Figure 2 we also show that I-DLA and SPARK behaves similarly, with all comparison methods exhibiting similar linear time growth. As expected, since the discrete gradient methods have no energy error, they also have zero variance up to round off error.

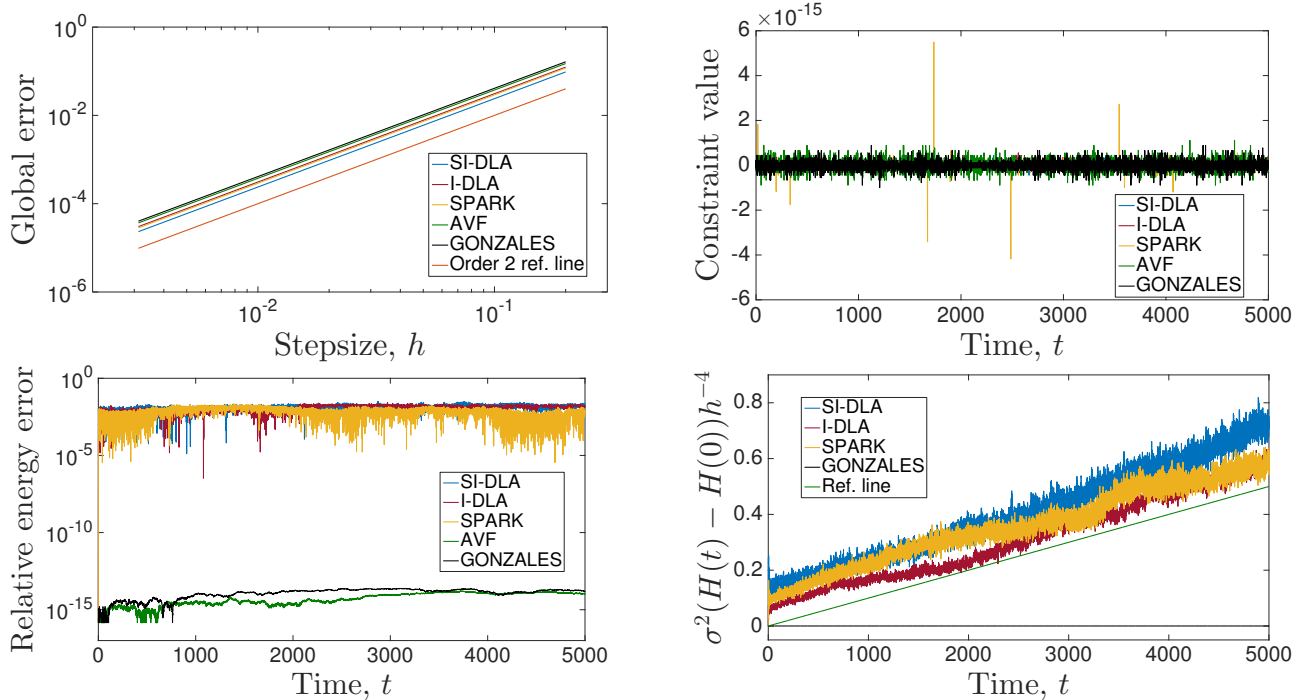


FIGURE 2. Comparison of the different methods for the fully chaotic system (19)-(20). **Top left:** Order plot, integrating up to $t = 10$. All methods are seen to be second order. **Top right:** Value of the left hand side of the constraint expression (20) for a sample trajectory with random initial conditions. The methods all respect the constraint up to machine precision. **Bottom left:** Relative energy error, i.e. $|H(t) - H(0)/H(0)|$, for the same trajectory. Only the discrete gradient methods conserve the energy up to machine precision. **Bottom right:** The variance of the energy errors $\sigma^2(H(t) - H(0))$ for 200 different initial conditions scaled by their expected h^4 dependence on the time step. The reference line $10^{-4}t$ is included for comparison. All comparison methods exhibit similar linear time growth in accordance with the reference line, while the discrete gradient method GONZALEZ has zero variance up to machine precision as expected. AVF is not shown since it was indistinguishable from GONZALEZ.

4.2. The Chaplygin sleigh. In this example we will see that the transformation of the systems into adapted coordinates can give rise to some additional numerical advantages apart from the possibility of achieving energy preservation.

The Chaplygin sleigh is a rigid body moving on a horizontal plane with three contact points, two of which slide freely without friction. The third one is a knife edge, which imposes the nonholonomic constraint of no motion perpendicular to the direction of the blade. The configuration space is $Q = SE(2)$, with coordinates (x_1, x_2, θ) . The coordinates (x_1, x_2) denote the contact point of the blade with the plane and θ the orientation of the blade. The Lagrangian is of kinetic type and if we assume that the center of mass lies in the line through the blade then it is given by

$$L = \frac{1}{2} \left((J + ma^2)\dot{\theta}^2 + m \left(\dot{x}_1^2 + \dot{x}_2^2 + 2a\dot{\theta}(-\dot{x}_1 \sin(\theta) + \dot{x}_2 \cos(\theta)) \right) \right),$$

where m denotes the mass of the body, J the moment of inertia relative to the center of mass and a the distance between the center of mass and the contact point of the blade. The matrix of the metric defining the kinetic Lagrangian is then given by

$$\begin{pmatrix} m & 0 & -ma \sin(\theta) \\ 0 & m & ma \cos(\theta) \\ -ma \sin(\theta) & ma \cos(\theta) & J + ma^2 \end{pmatrix}.$$

The nonholonomic constraint is $-\dot{x}_1 \sin(\theta) + \dot{x}_2 \cos(\theta) = 0$, which defines the non-integrable distribution

$$\mathcal{D} = \text{span} \left\{ \frac{\partial}{\partial \theta}, \cos(\theta) \frac{\partial}{\partial x_1} + \sin(\theta) \frac{\partial}{\partial x_2} \right\}.$$

For more details on this system, see [29].

In [13] there is a qualitative study of the DLA method when applied to the Chaplygin sleigh. More precisely, it is shown that the discrete momentum dynamics reproduces the same qualitative behaviour as the continuous momentum dynamics, as long as $|\theta' - \theta| < 2\pi$ and the momentum variable ρ_2 satisfies some bound. In the present example we examine the same issue using a discrete gradient method to the equations in adapted coordinates. We will obtain a bound on h but no bound on the momentum variables.

To derive the equations in adapted coordinates, we choose the following orthonormal basis adapted to \mathcal{D} and $\mathcal{D}^{\perp, g}$:

$$\begin{aligned} \mathcal{D} &= \text{span} \left\{ X_1 = \frac{1}{\sqrt{J + ma^2}} \frac{\partial}{\partial \theta}, X_2 = \frac{1}{\sqrt{m}} \left(\cos(\theta) \frac{\partial}{\partial x_1} + \sin(\theta) \frac{\partial}{\partial x_2} \right) \right\}, \\ \mathcal{D}^{\perp, g} &= \text{span} \left\{ X_3 = \frac{1}{\sqrt{\frac{(J + ma^2)^2}{ma^2} - (J + ma^2)}} \left(\frac{(J + ma^2)}{ma} \sin(\theta) \frac{\partial}{\partial x_1} - \frac{(J + ma^2)}{ma} \cos(\theta) \frac{\partial}{\partial x_2} + \frac{\partial}{\partial \theta} \right) \right\}, \end{aligned}$$

and denote the induced coordinates on TQ by $(x_1, x_2, \theta, y^1, y^2, y^3)$. In these coordinates the restricted Lagrangian $l : \mathcal{D} \rightarrow \mathbb{R}$ is given by $l(q^i, y^a) = \frac{1}{2}((y^1)^2 + (y^2)^2)$, and the nonholonomic constraint by $y^3 = 0$.

Since we have chosen an orthonormal basis we have the restricted Hamiltonian $\mathcal{H}(q^i, \rho_1, \rho_2) = \frac{1}{2} (\rho_1^2 + \rho_2^2)$, where $\rho_1 = \frac{\partial l}{\partial y^1} = y^1, \rho_2 = \frac{\partial l}{\partial y^2} = y^2$. Then

$$\Pi(\theta, \rho_1) = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{\cos(\theta)}{\sqrt{m}} \\ 0 & 0 & 0 & 0 & \frac{\sin(\theta)}{\sqrt{m}} \\ 0 & 0 & 0 & \frac{1}{\sqrt{J+ma^2}} & 0 \\ 0 & 0 & -\frac{1}{\sqrt{J+ma^2}} & 0 & -\frac{a\sqrt{m}}{J+ma^2}\rho_1 \\ -\frac{\cos(\theta)}{\sqrt{m}} & -\frac{\sin(\theta)}{\sqrt{m}} & 0 & \frac{a\sqrt{m}}{J+ma^2}\rho_1 & 0 \end{pmatrix}, \quad \nabla \mathcal{H}(\zeta) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \rho_1 \\ \rho_2 \end{pmatrix},$$

and the equations of motion (9) are for the position

$$\dot{x}_1 = \frac{\cos(\theta)}{\sqrt{m}}\rho_2, \quad \dot{x}_2 = \frac{\sin(\theta)}{\sqrt{m}}\rho_2, \quad \dot{\theta} = \frac{1}{\sqrt{J+ma^2}}\rho_1,$$

and for the momenta

$$(22) \quad \dot{\rho}_1 = -\frac{a\sqrt{m}}{J+ma^2}\rho_1\rho_2, \quad \dot{\rho}_2 = \frac{a\sqrt{m}}{J+ma^2}\rho_1^2.$$

The obtained equations are rather simple, since we have a quadratic vector field, a quadratic Hamiltonian and no constraints.

The mean value discrete gradient (11), the midpoint discrete gradient (12) and the coordinate increment discrete gradient (13) all coincide and give $\bar{\nabla} \mathcal{H}(\zeta, \zeta') = \left(0, 0, 0, \frac{\rho_1 + \rho'_1}{2}, \frac{\rho_2 + \rho'_2}{2}\right)^T$. As an approximation to the matrix Π we have chosen the midpoint value $\tilde{\Pi}(\zeta, \zeta') = \Pi\left(\frac{\zeta + \zeta'}{2}\right)$. Recalling Remark 3.2, the energy-preserving integrator (14) with any of these discrete gradients then collapses to the implicit midpoint rule, and is consequently given by

$$(23a) \quad x'_1 = x_1 + \frac{h}{2\sqrt{m}} \cos\left(\frac{\theta + \theta'}{2}\right) (\rho_2 + \rho'_2),$$

$$(23b) \quad x'_2 = x_2 + \frac{h}{2\sqrt{m}} \sin\left(\frac{\theta + \theta'}{2}\right) (\rho_2 + \rho'_2),$$

$$(23c) \quad \theta' = \theta + \frac{h}{2} \frac{1}{\sqrt{J+ma^2}} (\rho_1 + \rho'_1),$$

$$(23d) \quad \rho'_1 = \rho_1 - \frac{h}{4} \frac{a\sqrt{m}}{J+ma^2} (\rho_1 + \rho'_1) (\rho_2 + \rho'_2),$$

$$(23e) \quad \rho'_2 = \rho_2 + \frac{h}{4} \frac{a\sqrt{m}}{J+ma^2} (\rho_1 + \rho'_1)^2.$$

We will write the equations (23d) and (23e) as

$$F(\rho_1, \rho_2, \rho'_1, \rho'_2, h) := \rho'_1 - \rho_1 + \frac{h}{4} C_{12}^1 (\rho_1 + \rho'_1) (\rho_2 + \rho'_2) = 0,$$

$$G(\rho_1, \rho_2, \rho'_1, \rho'_2, h) := \rho'_2 - \rho_2 + \frac{h}{4} C_{21}^1 (\rho_1 + \rho'_1)^2 = 0.$$

Notice that $F(\rho_1, \rho_2, \rho_1, \rho_2, 0) = G(\rho_1, \rho_2, \rho_1, \rho_2, 0) = 0$ and compute

$$\left(\begin{array}{cc} \frac{\partial F}{\partial \rho'_1} & \frac{\partial F}{\partial \rho'_2} \\ \frac{\partial G}{\partial \rho'_1} & \frac{\partial G}{\partial \rho'_2} \end{array} \right) \bigg|_{(\rho_1, \rho_2, \rho_1, \rho_2, 0)} = \left(\begin{array}{cc} 1 + \frac{h}{4} C_{12}^1 (\rho_2 + \rho'_2) & \frac{h}{4} C_{12}^1 (\rho_1 + \rho'_1) \\ \frac{h}{2} C_{21}^1 (\rho_1 + \rho'_1) & 1 \end{array} \right) \bigg|_{(\rho_1, \rho_2, \rho_1, \rho_2, 0)} = \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right).$$

By the implicit function theorem we can write $\rho'_1 = f(\rho_1, \rho_2, h)$ and $\rho'_2 = g(\rho_1, \rho_2, h)$ in a neighbourhood of $(\rho_1, \rho_2, 0)$, with (ρ'_1, ρ'_2) also in a neighbourhood of (ρ_1, ρ_2) .

The continuous system has certain qualitative characteristics. Specifically, for the continuous system we have from (22), in the case $a \neq 0$, a one-dimensional manifold of equilibria $\{\rho_1 = 0\}$. These equilibria

are stable and asymptotically stable with respect to ρ_1 if $\rho_2 > 0$ and unstable if $\rho_2 < 0$. We will now study how the qualitative behaviour of (23d)-(23e) compares, as in [13].

Equilibria: If $h \neq 0$ then $F(\rho_1, \rho_2, \rho_1, \rho_2, h) = 0$ and $G(\rho_1, \rho_2, \rho_1, \rho_2, h) = 0$ imply $\rho_1 = 0$. Then the set $\{\rho_1 = 0\}$ is a one-dimensional manifold of equilibria.

Stability: Now we study the linearization of (f, g) at the equilibrium points $eq = (0, \rho_2, 0, \rho_2, h)$. Assuming that $\rho_2 \neq 0$ and $h < \left| \frac{2}{C_{12}^1 \rho_2} \right|$ we compute

$$\left(\begin{array}{cc} \frac{\partial f}{\partial \rho_1} & \frac{\partial f}{\partial \rho_2} \\ \frac{\partial g}{\partial \rho_1} & \frac{\partial g}{\partial \rho_2} \end{array} \right) \Big|_{eq} = \left(\begin{array}{cc} \frac{2-hC_{12}^1 \rho_2}{2+hC_{12}^1 \rho_2} & 0 \\ 0 & 1 \end{array} \right),$$

with eigenvalues

$$\lambda_1 = \frac{2-hC_{12}^1 \rho_2}{2+hC_{12}^1 \rho_2} = \frac{2(J+ma^2)-ha\sqrt{m}\rho_2}{2(J+ma^2)+ha\sqrt{m}\rho_2}, \quad \lambda_2 = 1.$$

Since $0 < h < \left| \frac{2(J+ma^2)}{a\sqrt{m}\rho_2} \right| = \left| \frac{2}{C_{12}^1 \rho_2} \right|$, we have $\lambda_1 > 0$, regardless of $\rho_2 \neq 0$. Further if $\rho_2 > 0$ then $\lambda_1 < 1$ and hence the equilibrium is stable and asymptotically stable with respect to ρ_1 . On the other hand if $\rho_2 < 0$ then $\lambda_1 > 1$ and hence the equilibrium is unstable. Therefore the proposed discrete method reproduces the same qualitative behaviour as the continuous system. This is not guaranteed when applying the midpoint rule to the Chaplygin sleigh system in the original coordinates.

Proposition 4.1. *The energy-preserving method (23d)-(23e) has a one-dimensional manifold of equilibria $\{\rho_1 = 0\}$. Assuming that $h < \left| \frac{2}{C_{12}^1 \rho_2} \right|$, the equilibria $(0, \rho_2)$ are stable and asymptotically stable with respect to ρ_1 if $\rho_2 > 0$ and are unstable if $\rho_2 < 0$.*

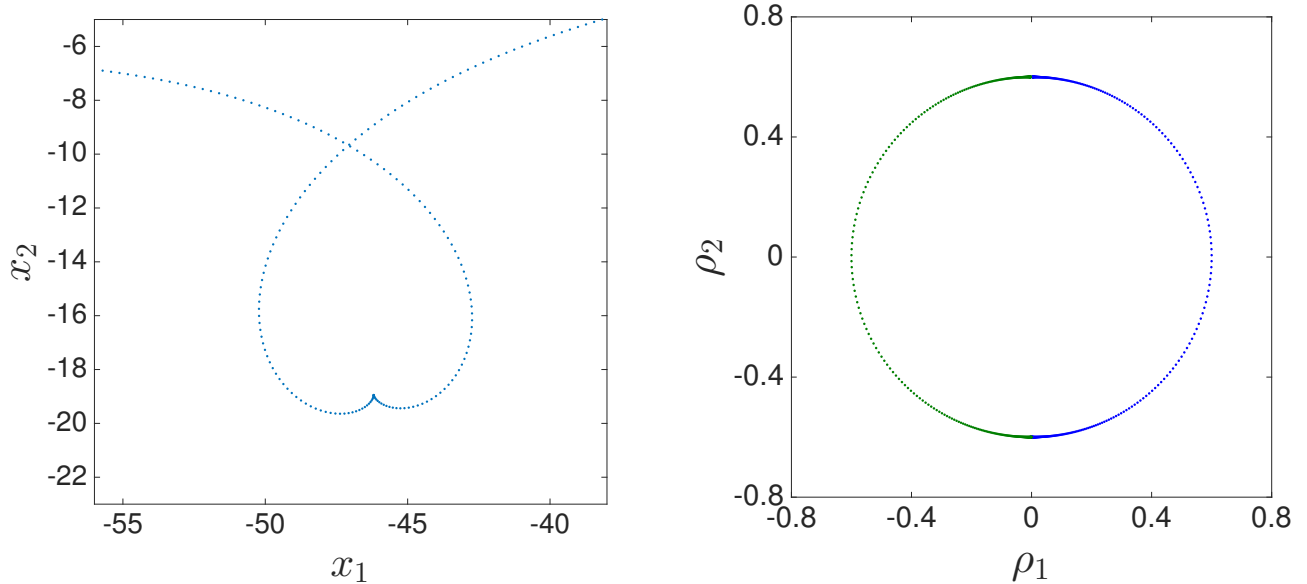


FIGURE 3. Integration results for the sleigh, using the method (23), with the parameters set to $J = 8$, $a = m = 1$, step-size $h = 0.5$ and initial values $x_1 = -5$, $x_2 = 0$, $\theta = 0.1$, $\rho_1 \in \{-0.001, 0.001\}$, $\rho_2 = -0.6$.

Left: A partial $x_1 x_2$ trajectory. **Right:** Two $\rho_1 \rho_2$ trajectories.

In Figure 3 we see an example of how the method exhibits correct behaviour by converging towards a stable equilibrium point when starting very close to an unstable one.

Remark 4.2. Similarly it is possible to show that any convergent Runge-Kutta method will give the correct behaviour for small enough h , when applied to the equations (22). For example, applying the explicit Euler method to these equations, we obtain the same conclusion as in Proposition 4.1 if we assume $h < \left| \frac{1}{C_{12}^1 \rho_2} \right|$. This confirms the fact that the illustrated good qualitative behaviour with respect to stability

of equilibria is more an effect of the choice of coordinates rather than of the choice of the method applied in those coordinates.

4.3. Euler-Poincaré-Suslov problem on $\mathfrak{so}(3)$. In this example we show that the approach that we have presented is also valid for nonholonomic systems defined on a Lie algebra (and more generally on a Lie algebroid [7]).

Let $\{e_1, e_2, e_3\}$ be a basis of the Lie algebra $\mathfrak{so}(3) \cong \mathbb{R}^3$ and denote the corresponding coordinates by $(\omega_1, \omega_2, \omega_3)$. Consider a kinetic Lagrangian on $\mathfrak{so}(3)$ defined by the matrix

$$(g_{ij}) = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{12} & I_{22} & I_{23} \\ I_{13} & I_{23} & I_{33} \end{pmatrix},$$

and introduce the nonholonomic constraints given by $\sum a_i \omega_i = 0$, where $a \in \mathfrak{so}(3)$ is a fixed element. We can choose the frame $\{e_1, e_2, e_3\}$ in such a way that $I_{12} = 0$ and $a = e_3$. Then the Lagrangian is given by

$$L = \frac{1}{2}(I_{11}\omega_1^2 + I_{22}\omega_2^2 + I_{33}\omega_3^2 + 2I_{13}\omega_1\omega_3 + 2I_{23}\omega_2\omega_3),$$

and the constraint reduces to $\omega_3 = 0$. This defines the distribution

$$\mathcal{D} = \text{span} \{X_1 := (1, 0, 0), X_2 := (0, 1, 0)\}.$$

Since the bracket on $\mathfrak{so}(3)$ is given by the cross product, it is immediate that \mathcal{D} is not involutive, and hence the constraint $\omega_3 = 0$ is nonholonomic. On the other hand,

$$\mathcal{D}^{\perp, g} = \text{span} \{X_3 := (I_{22}I_{13}, I_{11}I_{23}, -I_{11}I_{22})\}.$$

The Lie bracket of X_1 and X_2 is expressed in terms of the adapted basis X_1, X_2, X_3 as

$$[X_1, X_2] = (1, 0, 0) \times (0, 1, 0) = (0, 0, 1) = \frac{I_{13}}{I_{11}}X_1 + \frac{I_{23}}{I_{22}}X_2 - \frac{1}{I_{11}I_{22}}X_3.$$

Thus the nonvanishing structure constants of the projected bracket are

$$C_{12}^1 = \frac{I_{13}}{I_{11}} \quad \text{and} \quad C_{12}^2 = \frac{I_{23}}{I_{22}}.$$

If we denote by (y^1, y^2, y^3) the coordinates corresponding to the adapted basis $\{X_1, X_2, X_3\}$, the change of coordinates is given by

$$\omega_1 = y^1 + I_{22}I_{13}y^3, \quad \omega_2 = y^2 + I_{11}I_{23}y^3, \quad \omega_3 = -I_{11}I_{22}y^3.$$

Then the restricted Lagrangian becomes $l = \frac{1}{2}(I_{11}(y^1)^2 + I_{22}(y^2)^2)$ and the nonholonomic constraint is $y^3 = 0$.

In this example, since there are no (q^i) variables, the equations of motion (7a)-(7b) reduce to $\dot{\rho}_a = -C_{ab}^c \rho_c \frac{\partial \mathcal{H}}{\partial \rho_b}$, where $\mathcal{H} = \frac{1}{2} \left(\frac{1}{I_{11}} \rho_1^2 + \frac{1}{I_{22}} \rho_2^2 \right)$ and $\rho_i = \frac{\partial l}{\partial y^i} = I_{ii} y^i$, that is

$$\dot{\rho}_1 = -\frac{1}{I_{22}}(C_{12}^1 \rho_1 + C_{12}^2 \rho_2) \rho_2 \quad \text{and} \quad \dot{\rho}_2 = -\frac{1}{I_{11}}(C_{21}^1 \rho_1 + C_{21}^2 \rho_2) \rho_1.$$

In matrix form, using $C_{ab}^c = -C_{ba}^c$, we get

$$\begin{pmatrix} \dot{\rho}_1 \\ \dot{\rho}_2 \end{pmatrix} = - \begin{pmatrix} 0 & C_{12}^1 \rho_1 + C_{12}^2 \rho_2 \\ -(C_{12}^1 \rho_1 + C_{12}^2 \rho_2) & 0 \end{pmatrix} \begin{pmatrix} \frac{\rho_1}{I_{11}} \\ \frac{\rho_2}{I_{22}} \end{pmatrix}.$$

We apply the same discrete gradient method as in the previous example to get the integrator

$$\begin{pmatrix} \frac{\rho'_1 - \rho_1}{h} \\ \frac{\rho'_2 - \rho_2}{h} \end{pmatrix} = - \begin{pmatrix} 0 & C_{12}^1 \frac{\rho'_1 + \rho_1}{2} + C_{12}^2 \frac{\rho'_2 + \rho_2}{2} \\ - \left(C_{12}^1 \frac{\rho'_1 + \rho_1}{2} + C_{12}^2 \frac{\rho'_2 + \rho_2}{2} \right) & 0 \end{pmatrix} \begin{pmatrix} \frac{\rho'_1 + \rho_1}{2I_{11}} \\ \frac{\rho'_2 + \rho_2}{2I_{22}} \end{pmatrix}.$$

As mentioned previously the integrator is here equivalent to the implicit midpoint method.

4.4. Continuous gearbox driven by an asymmetric pendulum. In this final example we compare the performance of integrators applied directly to the formulation of the system in canonical coordinates. We consider a continuous gearbox driven by an asymmetric pendulum. This is a special case of the continuous gearbox system discussed in [27]. Here $Q = \mathbb{R}^3$ with Hamiltonian

$$(24a) \quad H(q^i, p_i) = \frac{1}{2} (p_1^2 + p_2^2 + p_3^2) - V(q^i),$$

$$(24b) \quad V(q^i) = \frac{1}{2} ((q^1)^2 + (q^2)^2) + \cos(q^3) - \frac{1}{5} \sin(2q^3).$$

The single nonholonomic linear velocity constraint is

$$\dot{q}^1 + \sin(q^3)\dot{q}^2 = p_1 + \sin(q^3)p_2 = 0,$$

since clearly $p_i = \dot{q}^i$, $1 \leq i \leq 3$ because $(g_{ij}) = I_3$.

Again comparing with the semi implicit reversible DLA variational integrator (SI-DLA) proposed in [24], we consider from Section 3.3 the initial method (15) using the Gonzalez midpoint discrete gradient (Gonzalez-F) and the canonical coordinate implementation of the Gonzalez midpoint discrete gradient method for the reduced system (Gonzalez-R). For the methods we compare in Figure 4 the relative energy error, i.e. $|(H - H_0)/H_0|$, and the constraint error, i.e. $|p_1 + \sin(q^3)p_2|$ for a long time simulation $t \in [0, 50000]$ with random initial values chosen to ensure non-periodic behaviour. H_0 is the initial energy.

For SI-DLA we observe an exponential growth in the energy error, while both Gonzalez methods preserve the energy to machine precision. SI-DLA and GONZALEZ-R both preserve the nonholonomic constraint to machine precision. However the energy blow-up of SI-DLA gives a corresponding increase in the round off error and thus the constraint error over time. GONZALEZ-F does not respect the nonholonomic constraint. The results are thus as expected.

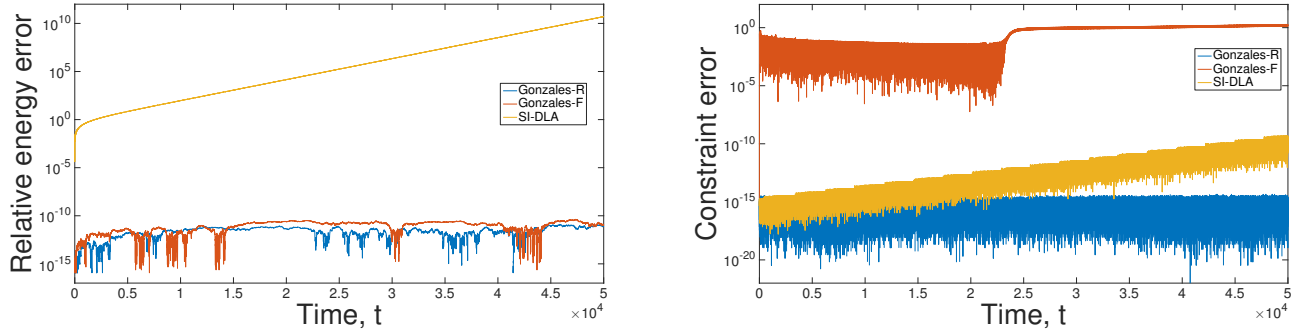


FIGURE 4. Integration results for methods GONZALEZ-R, GONZALEZ-F and SI-DLA, applied to the system (24) with random initial values, $h = 0.1$, and $t \in [0, 50000]$. **Left:** Relative energy error, i.e. $|H(t) - H(0)/H(0)|$. **Right** The error in the nonholonomic constraint, i.e. $|p_1 + \sin(q^3)p_2|$.

5. FUTURE WORK

In this paper we have not addressed the case when Q is a differentiable manifold. In a future paper, we will propose to adapt the discrete gradient approach taking the geometry of the configuration space into account, see for instance the methods in reference [5]. In order to adapt the ideas in [5] to a general differentiable manifold Q , we will need to introduce a finite difference map or retraction map $\Phi_h : U \subset \mathcal{D}^* \times \mathcal{D}^* \rightarrow T\mathcal{D}^*$ (see [24]) from a finite difference map initially defined on Q . In this case we will define a discrete gradient as a map $\bar{\nabla}\mathcal{H} : \mathcal{D}^* \times \mathcal{D}^* \rightarrow T^*\mathcal{D}^*$ verifying similar properties to Definition 3.1 (see [5]).

$$\begin{array}{ccc} \mathcal{D}^* \times \mathcal{D}^* & \xrightarrow{\bar{\nabla}\mathcal{H}} & T^*\mathcal{D}^* \\ \downarrow \Phi_h & & \downarrow \pi_{\mathcal{D}^*} \\ T\mathcal{D}^* & \xrightarrow{\tau_{\mathcal{D}^*}} & \mathcal{D}^* \end{array}$$

In this case, an energy preserving integrator for Equation (9) would be

$$\Phi_h(\zeta, \zeta') = \Pi(\bar{\zeta})\bar{\nabla}\mathcal{H}(\zeta, \zeta')$$

with $\bar{\zeta} = \tau_{\mathcal{D}}^*(\Phi_h(\zeta, \zeta'))$. We will explore this possibility in a future paper since in many examples of nonholonomic systems the configuration space is a nonlinear space such as, for instance, a Lie group G .

Moreover, it would be interesting to compare the discrete gradient method approach introduced in this paper with other methods designed for nonholonomic systems. For instance, the Chaplygin case is given by a Lagrangian system with forces on the tangent space of a reduced space and then it is possible to use directly discrete variational integrators based on forced Lagrangian systems (see [8, 9]). Other interesting possibilities to compare our methods with are variational integrators from Hamiltonizable nonholonomic systems [14] or the geometric nonholonomic integrator [15].

ACKNOWLEDGMENTS

This work has been partially supported by MINECO (Spain) MTM2013-42870-P, MTM2015-69124-REDT, the ICMAT Severo Ochoa project SEV-2015-0554 and the Nils-Abel project 010-ABEL-CM-2014ANILS. It has also received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 691070. MFP has been financially supported by a FPU scholarship from MECED.

REFERENCES

- [1] V. I. Arnol'd, V. V. Kozlov, and A. I. Neishtadt. *Dynamical systems. III*, volume 3 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 1988. Translated from the Russian by A. Iacob. 4
- [2] P. Balseiro, M. de León, J. C. Marrero, and D. Martín de Diego. The ubiquity of the symplectic Hamiltonian equations in mechanics. *J. Geom. Mech.*, 1(1):1–34, 2009. 7
- [3] A.M. Bloch. *Nonholonomic mechanics and control*, volume 24 of *Interdisciplinary Applied Mathematics*. Springer Verlag, 2003. 1, 2, 3
- [4] A.M. Bloch, J.E. Krishnaprasad, J.E. Marsden, and R. Murray. Nonholonomic mechanical systems with symmetry. *Arch. Rat. Mech. An.*, 136:21–99, 1996. 1, 2
- [5] Elena Celledoni and Brynjulf Owren. Preserving first integrals with symmetric Lie group methods. *DCDS-A*, 34(3):977–990, 2014. 2, 8, 19
- [6] D Cohen and E. Hairer. Linear energy-preserving integrators for poisson systems. *BIT Numer Math*, 51:91–101, 2011. 9
- [7] J. Cortés, M. de León, J. C. Marrero, and E. Martínez. Nonholonomic Lagrangian systems on Lie algebroids. *Discrete Contin. Dyn. Syst.*, 24(2):213–271, 2009. 18
- [8] J. Cortés and S. Martínez. Non-holonomic integrators. *Nonlinearity*, 14(5):1365–1392, 2001. 1, 20
- [9] J. Cortés Monforte. *Geometric, control and numerical aspects of nonholonomic systems*, volume 1793 of *Lecture Notes in Mathematics*. Springer, 2002. 1, 2, 20
- [10] M. de León, J. C. Marrero, and D. Martín de Diego. Linear almost Poisson structures and Hamilton-Jacobi equation. Applications to nonholonomic mechanics. *J. Geom. Mech.*, 2(2):159–198, 2010. 7
- [11] M. de León, D. Martín de Diego, and A. Santamaria-Merino. Geometric integrators and nonholonomic mechanics. *J. Math. Phys.*, 45:1042–1064, 2004. 1
- [12] Y. N. Fedorov and D. V. Zenkov. Discrete nonholonomic LL systems on Lie groups. *Nonlinearity*, 18(5):2211–2241, 2005. 1
- [13] Y. N. Fedorov and D. V. Zenkov. Dynamics of the discrete Chaplygin sleigh. *Discrete Contin. Dyn. Syst.*, (suppl.):258–267, 2005. 15, 17
- [14] Oscar E. Fernandez, Anthony M. Bloch, and Peter J. Olver. Variational integrators from Hamiltonizable nonholonomic systems. *J. Geom. Mech.*, 4(2):137–163, 2012. 20
- [15] S. Ferraro, D. Iglesias, and D. Martín de Diego. Momentum and energy preserving integrators for nonholonomic dynamics. *Nonlinearity*, 21(8):1911–1928, 2008. 1, 20
- [16] G.H. Golub and C.F. Van Loan. *Matrix computations*. The John Hopkins University Press, Baltimore, Maryland and London, England, third edition, 1996. 11
- [17] O. Gonzalez. Time integration and discrete Hamiltonian systems. *J. Nonlinear Sci.*, 6(5):449–467, 1996. 2, 8, 9
- [18] J. Grabowski, M. de León, J. C. Marrero, and D. Martín de Diego. Nonholonomic constraints: a new viewpoint. *J. Math. Phys.*, 50(1):013520, 17, 2009. 6, 7
- [19] E Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2010. Structure-preserving algorithms for ordinary differential equations, Reprint of the second (2006) edition. 1, 9
- [20] A. Harten, P. D. Lax, and B. van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Rev.*, 25(1):35–61, 1983. 9

- [21] T. Itoh and K. Abe. Hamiltonian-conserving discrete canonical equations based on variational difference quotients. *J. Comput. Phys.*, 76(1):85–102, 1988. 9
- [22] L. O. Jay. Solution of index 2 implicit differential-algebraic equations by Lobatto Runge-Kutta methods. *BIT*, 43(1):93–106, 2003. 14
- [23] M. Kobilarov, D. Martín de Diego, and S. Ferraro. Simulating nonholonomic dynamics. *Bol. Soc. Esp. Mat. Apl. SEMA*, (50):61–81, 2010. 1
- [24] R. McLachlan and M. Perlmutter. Integrators for nonholonomic mechanical systems. *J. Nonlinear Sci.*, 16(4):283–328, 2006. 1, 11, 12, 14, 19
- [25] R. I. McLachlan and G. R. W. Quispel. Geometric integrators for ODEs. *J. Phys. A*, 39(19):5251–5285, 2006. 1
- [26] R. I. McLachlan, G. R. W. Quispel, and N. Robidoux. Geometric integration using discrete gradients. *R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci.*, 357(1754):1021–1045, 1999. 2, 8, 9
- [27] K. Modin and O. Verdier. Integrability of nonholonomically coupled oscillators. *Discrete Contin. Dyn. Syst.*, 34:1121–1130, 2014. 19
- [28] R. M. Murray and S. S. Sastry. Nonholonomic motion planning: steering using sinusoids. *IEEE Trans. Automat. Control*, 38(5):700–716, 1993. 1
- [29] J.I. Neimark and N.A. Fufaev. *Dynamics of nonholonomic systems*. Translations of the American Mathematical Society, Providence, Rhode Island, 1972. 1, 6, 15
- [30] R. A. Norton, D. I. McLaren, G. R. W. Quispel, A. Stern, and A. Zanna. Projection methods and discrete gradient methods for preserving first integrals of ODEs. *Discrete Contin. Dyn. Syst.*, 35(5):2079–2098, 2015. 2
- [31] G. R. W. Quispel and H. W. Capel. Solving ODEs numerically while preserving a first integral. *Phys. Lett. A*, 218(3-6):223–228, 1996. 2
- [32] G. R. W. Quispel and G. S. Turner. Discrete gradient methods for solving ODEs numerically while preserving a first integral. *J. Phys. A*, 29(13):L341–L349, 1996. 2
- [33] J. M. Sanz-Serna and M. P. Calvo. *Numerical Hamiltonian problems*, volume 7 of *Applied Mathematics and Mathematical Computation*. Chapman & Hall, London, 1994. 1
- [34] A. J. van der Schaft and B. M. Maschke. On the Hamiltonian formulation of nonholonomic mechanical systems. *Rep. Math. Phys.*, 34(2):225–233, 1994. 1, 7

¹ DEPARTMENT OF MATHEMATICAL SCIENCES, NTNU, 7491 TRONDHEIM, NORWAY

E-mail address, Elena Celledoni: `elenac@math.ntnu.no`

E-mail address, Eirik Hoel Høiseth: `eirik.hoiseth@math.ntnu.no`

² INSTITUTO DE CIENCIAS MATEMÁTICAS (CSIC-UAM-UC3M-UCM), CALLE NICOLÁS CABRERA 15, CAMPUS UAM, CANTOBLANCO, MADRID, 28049, SPAIN

E-mail address, Marta Farré Puiggalí: `marta.farre@icmat.es`

E-mail address, David Martín de Diego: `david.martin@icmat.es`